Peptide Selection for the Quantification of P-III-NP in Human Serum by Mass Spectrometry

Journal:	*Rapid Communications in Mass Spectrometry*
Manuscript ID	RCM-17-0309.R1
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	n/a
Keywords:	procollagen III amino-terminal propeptide, P-III-NP, post-translational modification, in silico digestion, tryptic digest, doping control

Complete List of Authors: Moncrieffe, Danielle; King's College London, Department of Analytical, Environmental and Forensic Sciences
Parkin, Mark; King's College London, Department of Environmental, Analytical and Forensic Science
Cowan, David; King's College London, Department of Analytical, Environmental and Forensic Sciences

Abstract: RATIONALE: Procollagen III amino-terminal propeptide (P-III-NP) is monitored using immunoassays as a biomarker for growth hormone administration in human doping control and in clinical diagnostics. These immunoassays have known drawbacks and research is ongoing to develop a mass spectrometric (MS) method. A lack of traceable reference material, presence of post translational modifications (PTMs), and small blood concentration complicate the development of targeted analytical methods for P-III-NP quantification.
METHODS: Tryptic digest products of P-III-NP were assessed by liquid chromatography-MS (LC/MS). *In silico* digestion was used to predict P-III-NP peptides for MS analysis; however these excluded PTMs. With a priori knowledge of PTMs, we associated experimental P-III-NP peptides with those derived by in silico digestion. Synthesized P-III-NP peptides, hT1 (human) and T5 (human/bovine), were used to develop sensitive micro- and nano-flow LC/MS methods to analyse P-III-NP, originating from human serum, semi-quantitatively.
RESULTS: P-III-NP peptides, T1 and T5, were identified using high resolution accurate MS (HRAMS). PTMs modified the mass of observed peptides. N-terminal pyroglutamation (pE) in T1 and several hydroxylated prolines (hP) in T5 (G-X-hP motif) were observed. With PTM, hT1 and T5 were observed in a digest of immuno-captured P III NP by LC/MS. Using a semi quantitative approach, hP-III-NP at a basal concentration of 2 ng/mL (50 pmol) could be estimated from a 200 µL sample volume.
CONCLUSIONS: Consideration of PTMs is needed to identify P-III-NP peptides produced by digestion with trypsin. The information presented here now gives the most appropriate peptide sequences for synthesizing suitable reference material required for quantification of human P-III-NP in blood and evidences methodology that is sufficiently sensitive to develop a

	quantitative method.

Peptide Selection for the Quantification of P-III-NP in Human Serum by Mass Spectrometry

Danielle Moncrieffe, Mark C. Parkin, David Cowan*

Drug Control Centre, King's Forensics, Department of Analytical, Environmental and Forensic Sciences, King's College London, London, UK.

RATIONALE: Procollagen III amino-terminal propeptide (P-III-NP) is currently monitored in human doping control as a biomarker for growth hormone administration and also in clinical diagnostics using immunoassays. Drawbacks to this approach have been highlighted, and research is ongoing to develop a mass spectrometric (MS) method to complement these methods. However, a lack of traceable reference material, presence of post translational modifications (PTMs), and small blood concentration implicate the development of targeted analytical methods for P-III-NP quantification.

METHODS: Tryptic digest products of P-III-NP were assessed by liquid chromatography-MS (LC/MS). *In silico* digestion was used to predict P-III-NP peptides for MS analysis; however these excluded PTMs. With *a priori* knowledge of PTMs, we associated experimental P-III-NP peptides with those derived by *in silico* digestion. Synthesized P-III-NP peptides, hT1 (human) and T5 (human/bovine), were used to develop sensitive micro- and nano-flow LC MS methods to analyse P-III-NP, originating from human serum, semi-quantitatively.

RESULTS: P-III-NP peptides, T1 and T5, were identified using high resolution accurate MS (HRAMS). PTMs modified the mass of observed peptides. N-terminal pyroglutamation (*p*E) in T1 and several hydroxylated prolines (*h*P) in T5 (G-X-*h*P motif) were observed. With PTM, *h*T1 and T5, were observed in a digest of immuno-captured P-III-NP by LC/MS. Using a semi-quantitative approach, *h*P-III-NP at basal concentrations of 2 ng/mL (50 pmol) could be estimated from a 200 μL sample volume.

CONCLUSIONS: Consideration of PTMs is needed to identify P-III-NP peptides produced by digestion with trypsin. The information presented here now gives the most appropriate peptide sequences for synthesizing suitable reference material required for quantification of human P-III-NP in blood and evidences methodology that is sufficiently sensitive to develop a quantitative method.

Keywords: procollagen III amino-terminal propeptide; post-translational modification; *in silico* digestion; tryptic digest; doping control

*Corresponding author: david.a.cowan@kcl.ac.uk

Introduction

Type III collagen is the second most abundant collagen (varying in abundance with age) in humans; it is typically found with type I collagen and is present in tissues with elastic properties such as skin, blood vessels and various internal organs[1]. Type III collagen is formed from a precursor pro-collagen unit in the cell, which includes amino- and carboxyl-terminal extension propeptides (denoted P-III-NP and P-III-CP, respectively) [1, 2].

P-III-NP is a clinical biomarker used to monitor chronic active hepatitis, liver fibrosis and cirrhosis [3-6]. Elevated levels of P-III-NP have been associated with hormonally induced growth [2, 7], making P-III-NP one of two biomarkers used to detect rhGH (recombinant human growth hormone) doping in sport [8]; the other being insulin-like growth factor I (IGF-I). These compounds have little diurnal variation and are largely unaffected by exercise or sex, but respond to hGH administration. Also, body mass index and race contribute little to the variability of these markers [8]. For humans, blood P-III-NP levels are inversely proportional to age, with maximum levels observed during pubertal years. Sex however affects P-III-NP blood concentrations, men (average conc. 5.4 ±2.3 ng/mL) having a slightly bigger natural concentration than women (average conc.5.1 ±1.5 ng/mL)[9].

P-III-NP consists of three identical 130 amino acid pro α1- chains, connected by disulphide cysteine bonds to create a unique pro-collagen subunit with three distinct domains, Col 1-3 [1, 10]. The first domain, Col 1, has a globular structure resulting from its richness in acidic amino acids and cysteine. There are five intra-molecular disulphide bonds within each pro α1-chain forming this domain, causing it to fold into several loops to expose more hydrophilic amino acids increasing interaction with water [10-12]. The second domain, Col 3, is triple helical, consisting of the tri-peptide repeat G-X-Y, with X and Y most often being a proline residue. The high proline and glycine (G) content within this region results in a left-handed super-helical twist of the three pro α1 chains to form a "collagen-like" conformation. The six inter-molecular disulphide cysteine bonds in the telopeptide domain, Col 2, which can be found at the C-terminus of the peptide, ensure that the three chains are close together facilitating the twisted conformation[10, 11].

In doping control, measurements of P-III-NP have relied on the use of radioactive immunoassays (RIA), RIA-gnost® (CisBio Bioassays) and UniQ® P-III-NP RIA (Orion Diagnostica) and the non-radioactive Advia Centaur® immunoassay platform (Siemens Healthcare Diagnostics Inc.) [1, 13]. In the absence of international reference material for human P-III-NP, these immunoassays use bovine P-III-NP (97 % sequence homology with human) as reference; thus, relying on the cross- reactivity of epitopes for human and bovine material. More recently an enzyme linked immunosorbent assay (ELISA) (CisBio Bioassays) has been developed; however, it is still to be evaluated by anti-doping laboratories.

World Anti-doping Agency (WADA), the body that accredits laboratories to undertake human anti-doping analysis, stipulates that confirmation of elevated P-III-NP requires the use of two independent immunoassays recognising different epitopes (see WADA International Standard for Laboratories version 9.0, 2016 paragraph 5.2.4.3.1.3). This makes laboratories reliant on the use of RIAs, needing licensing for handling radioactive material. Apart from

the difficulties related to handling radioisotopes, other limitations are associated with these RIAs including their lack of selectivity for the human homologue, lack of harmonisation of units of measurement and epitopes used between methods, and possible change or withdrawal of assays by their manufacturers, as has occurred previously [14, 15].Furthermore, the decision limits distinguishing negative and positive results is method specific and need modification whenever an assay is changed. The second biomarker used to evidence human growth hormone administration is IGF-I, a 7.5 kDa protein with circulating concentrations typically in excess of 100 ng/mL [15]. With the acceptance by WADA of a validated mass spectrometry (MS) method for this protein, there is now more standardisation of IGF-I measurement. Thus, to overcome the limitations described, a MS method for the analysis of P-III-NP seems timely to complement or replace the already existing RIAs.

Absence of a traceable international reference material for human P-III-NP (*h*P-III-NP) makes the development of specific methods reliant on endogenous sources. The small serum concentration (25-125 pM) and medium protein size of P-III-NP (~42 kDa) [15], make it difficult to obtain adequate MS sensitivity from intact protein analysis and thus a digest approach seems appropriate. Digestion of proteins with the endoprotease trypsin usually yield peptide fragments for MS analysis with favourable ionisation properties [16]. Unique digestion peptide fragments of P-III-NP obtained from *in silico* digestion were searched against protein databases and selected for MS analysis. Although post-translational modifications (PTMs) of proteins are not uncommon, for P-III-NP, modifications are not present in the *in silico* generated reference peptide sequences. Based on the structure and amino acid composition of P-III-NP, PTMs can be predicted, but require confirmation; these affect the observed mass to charge ratio (*m/z*) of the ions obtained from the peptides selected from *in silico* digestion. PTMs across P-III-NP include pyroglutamation in Col 1 [11], hydroxylation of proline residues in the "collagen-like" Col 3 domain, de-amidation of susceptible asparagine residues across the protein, and possible N-linked glycosylation in Col 2 [11, 17].

Here we confirm the presence, in human serum, of the selected *in silico* P-III-NP peptides and their PTMs, for the first time. Trypsin digested bovine P-III-NP (*b*P-III-NP) was analysed by high-resolution accurate mass spectrometry (HRAMS) to identify peptides. Once identified, pure human variants of these peptides were synthesized and used to develop sensitive nano- and micro-flow liquid chromatography mass spectrometry (LC/MS) methods capable of identifying trypsin digested endogenous *h*P-III-NP separated from a pooled serum sample. This work highlights suitable tryptic peptides to enable the development of reliable quantification methods, once the necessary standardised reference material (native and heavy labelled) for these proteins or intact P-III-NP are available.

Materials and methods

All water used was purified using an Elga Purelab® flex water purification unit and had ~18 MΩ.cm resistivity. Sequencing grade modified trypsin (porcine) was purchased from Promega (Wisconsin, USA). Dithiothreitol (DTT), electrophoresis purity reagent, was purchased from Bio-Rad Laboratories (Hertfordshire, UK). Iodoacetamide (IDA, ≥99 %, HPLC), ammonium bicarbonate (≥99 %, Reagent Plus®), trifluoroacetic acid (99 %, Reagent Plus®), tetracosactide (adrenocorticotropic hormone fragment 1-24 (human, rat) ≥97 %) and

albumin solution human (HSA, 30 % in 0.85 % sodium chloride, protease free), were purchased from Sigma Aldrich (St. Louis, MO, USA). Analar NormaPur formic acid (99-100 %) was purchased from VWR Chemicals (Fontenay-sous-Bois, France). Acetonitrile (HPLC grade) and acetic acid glacial (Analytical Reagent grade) were purchased from Fisher Scientific (Loughborough, UK).

In silico digestion

Reference sequences for human (UniProt accession no. P02461(24-153)) and bovine (UniProt accession no.Q08E14 (24-153)) P-III-NP were accessed from UniProtKB (http://www.uniprot.org/uniprot/P02461, last accessed September 12, 2017) [18]. Theoretical digestion of these proteins with trypsin (no missed cleavage allowed) was done using the ExPasy-Peptide Cutter [19] and USCF Protein Prospector [20]. Digest products were assessed within UniProtKB using the basic local alignment tool (BLAST, http://www.uniprot.org/blast/, last accessed September 12 2017) to identify the uniqueness of the peptide fragments to P-III-NP.

Bovine P-III-NP sample

A *b*P-III-NP sample (~10 µg/mL, unspecified purity) in human serum albumin (HSA, 0.1 %) was kindly donated by the Institute of Bioanalytics, L.L.C (Branford, USA). This was assessed using UniQ® P-III-NP RIA (Orion Diagnostica, Espoo, Finland) to obtain an approximate concentration. HSA (0.1 %) in phosphate buffered saline (PBS, 0.01 M) both purchased from Sigma Aldrich (St. Louis, MO, USA) was used as negative control.

Trypsin digestion and LC/HRAMS analysis of bovine P-III-NP

P-III-NP and negative control (0.1 % HSA in 0.01 M phosphate buffered saline) samples were treated as follows. Each sample (100 µL) was evaporated to dryness and reconstituted in ammonium bicarbonate (50 µL, 100 mM) in a LoBind Eppendorf® tube (Sigma Aldrich, Hamburg, Germany). DTT in 100 mM ammonium bicarbonate (5 µL, 100 mM) was added to the samples, which were incubated at 60 °C with mixing in an Eppendorf Thermomixer® at 900 rpm for 10 min. Samples were left to incubate at 60 °C for a further 50 min without mixing. After cooling, freshly prepared IDA in 100 mM ammonium bicarbonate (5 µL, 250 mM) was added to the samples and vortexed before leaving in the dark for 30 min. Sequencing grade modified trypsin (4 µg) was added to the samples, which were incubated at 37 °C overnight with mixing at 1400 rpm. To stop trypsin activity, formic acid (10 µL, 2 %) was added to each sample before evaporating to dryness and reconstituting in mobile phase (50 µL, 0.3 % formic acid, 5 % acetonitrile in water). Samples were then transferred to a suitable autosampler vial and analysed by LC/HRAMS.

Using a Thermo Fisher™ Dionex™ UltiMate™ 3000 UHPLC (Ultra High Performance Liquid Chromatography) system (Waltham, MA, USA), gradient separation of each sample (10 µL) was achieved on a Waters Acquity® Ultra Performance LC BEH C18 column (130Å, 1.7 µm, 2.1 x 50 mm. Milford, MA, USA) with mobile phase A, 0.3 % formic acid in water, and B, 0.3 % formic acid in acetonitrile. Using a flow rate of 0.3 mL/min, mobile phase B was held at 5 % for 3 min, then increased to 90 % for 5 min before returning to 5 % for a 2 min column re-equilibration. MS data was acquired using a Thermo Fisher™ Q-Exactive™ HRAMS system with HESI II (heated ESI) ion source operating in positive mode at a

resolving power of 140,000 FWHM. Three separate experiments were conducted in positive ionisation mode: a full scan HRAMS, a targeted tandem mass spectrometry-high resolution mass spectrometry (MS/HRAMS), and a data dependent MS/HRAMS.

The ESI sheath and auxiliary nitrogen gas flows were set to 70 and 13 arbitrary units respectively, the sweep gas was set to zero. The capillary and heater temperature was 300 °C and 325 °C respectively. The spray voltage was 3.75 kV. Nitrogen was used as the source and collision gas. For full scan HRAMS experiments, an AGC target of 1×10^6 was used with a maximum injection time of 100 ms. Appropriate scan ranges were selected for initial HRAMS experiments to observe precursor ions with single, double or triple charges. MSHRAMS analyses were run using appropriate *m/z* values for the precursor ion (±1 u) where necessary. Top 10 and targeted data dependent analysis was performed for the T1 doubly charged ion, with an AGC target of 1.5×10^4.

LC/HRAMS data was reviewed on Thermo Scientific™ Xcalibur™ 2.2 SP 1.48 Qual Browser software; extracted ion chromatograms were produced with a tolerance of ±5 ppm. Data-dependent results were checked against the Matrix Science Mascot server (version 2.2.06) searching against all taxonomic entries in the UniProtKB_sprot_130200. Allowing for a peptide tolerance of 5 ppm with 3 missed cleavages, with carbamidomethylation as a fixed modification oxidation of proline, N-terminal pyroglutamation and asparagine deamidation as variable modifications.

Trypsin derived P-III-NP peptides

Authentic PTM *h*P-III-NP peptides, *h*T1 (***Q-***$_{NH3}$QEAVDGGCSHLGQSYADR) and T5 (GD***P-***$_{OH}$GP***P-***$_{OH}$GI***P-***$_{OH}$GR) were both synthesized and purchased from Thermo Fisher (Life Technologies limited, Paisley, UK).

Reduction of disulphide bridges in *h*T1 (100 µL, 1 µg/mL in 100 mM ammonium bicarbonate) was achieved by adding DTT in 100 mM ammonium bicarbonate (5 µL, 100mM) and incubating at 60 ˚C with mixing at 900 rpm for 10 min then, without mixing, for an additional 50 min. Disulphide sites were blocked by alkylating with IDA in 100 mM ammonium bicarbonate (5 µL, 250 mM) at room temperature for 30 min.

Pure peptide standards for *h*T1 (reduced and alkylated) and T5 were prepared in mobile phase (50 µL, 0.3 % formic acid, 5 % acetonitrile in water) at approximate concentrations: 5 pM, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM and 1 nM. Standards were analysed by micro- and nano-flow LC/MS to establish an approximate limit of detection (LOD); where the signal to noise is greater than 3:1 for all diagnostic ions.

Nano- and micro-LC/MS methods

Micro-flow and nano-flow LC/MS were performed on a Waters nanoAcquity® Ultra Performance LC® coupled to a Waters Xevo™ TQ-S MS in positive MS mode. For micro-flow, the mass spectrometer was interfaced with a Waters narrow gauge ESI needle (60 µm I.D.) on a conventional ion source. A low dead volume fused silica capillary (25 µm I.D) was used from the outlet of the column to the inlet of the ESI needle. The capillary voltage set to 2.22 kV and the cone voltage was set to 22 V with an offset of 27 V. The source and

desolvation temperature were 150 ˚C and 350 ˚C, respectively. The cone and desolvation gas flow were 145 and 450 L/h. Collision (argon) and nebuliser (nitrogen) gas flow was 0.17 mL/min and 7 bar, respectively. An MS mode collision energy of 4 eV was applied. For nano-flow, the mass spectrometer was interfaced with a nano-ESI emitter (30 µm I.D. New Objective, Woburn, MA, USA) using a Waters nano-ESI Z-Spray source (Waters, USA), with the capillary voltage set to 1.75 kV. The cone voltage was set to 45 V with an offset of 30 V. The source temperature was set to 83 ˚C, with a collision gas flow of 0.15 mL/min and a nebuliser gas flow of 7 bar. MS mode collision energy of 10 eV was applied.

Micro-flow LC gradient separation of each sample (1 µL) was achieved on an Acquity® M-Class BEH 130 Å (C18, 1.7 µm, 300 µm x 150 mm) column at 45 ˚C with mobile phase A, 0.3 % formic acid in water, and B, 0.3 % formic acid in acetonitrile and a flow rate of 8 µL/min. Mobile phase B was increased from 5 % to 60 % over 20 min, this was then increased to 90 % in 1.5 min and held for 30 sec, after which it was immediately returned to 5 % and held for 3 min to re-equilibrate the column.

For nano-flow LC, mobile phase A, 0.3 % formic acid in water, and B, 0.3 % formic acid in acetonitrile was used. Samples (5 µL) were trapped on an Acquity® UPLC® Symmetry C18 nanoAcquity® 10K 2GV/M trapping column (100 Å, 5 µm, 180 µm x 20 mm) for 5 min at flow 10 µL/min with 1 % mobile phase B. The trapped analytes were then loaded onto an Acquity® UPLC® M-Class Peptide BEH C18 column (130 Å, 17 µm, 75 µm x 150 mm) at 45 ˚C for separation by gradient elution. With a flow of 0.4 µL/min, mobile phase B was held at 5 % for 5 min, then increased to 9 % for 25 min and then 95 % for 2 min, before returning to 5 % and held for 8 min to re-equilibrate the column.

Selected reaction monitoring for T1 and T5 were acquired for the transitions shown in Table 1. LC/MS data was analysed using MassLynx Software version 4.1 (Waters, Milford, MA, USA).

Extraction of P-III-NP from human serum for LC/MS

Pooled human serum (100 µL) and HSA control (100 µL, 50 mg/mL) samples were brought to room temperature, before placing into a P3NP ELISA (CisBio Bioassays, Gif-sur-Yvette, France) microtitre plate containing immobilised anti-P-III-NP mouse monoclonal antibody. Samples were incubated overnight at 4 ˚C with shaking at 700 rpm. Liquid was then dispelled from the wells, which were then washed three times with PBS containing 0.3 % Tween® 20 (300 µL). Bound P-III-NP was removed from the wells by incubating with 15 mg/mL tetracosactide, 33 % acetonitrile, 0.4 % TFA solution (100 µL) at 60 ˚C for 5 min while shaking at 900 rpm. The samples were then transferred to LoBind Eppendorf® tubes and digested with trypsin, before analysing by micro- and nano-flow LC/MS. Samples were compared to a T5 standard (100 pM) to determine their concentrations semi-quantitatively.

Results and Discussion

Theoretical digestion of P-III-NP with trypsin gave four unique *in silico*-derived peptides, T1, T2-4, T5 and T6, shown in Table 2. Of these peptides, T2-4, contains two tryptic no cleavage sites, at ^{21}V-W-K-P-E and ^{81}P-T-R-P-P, making this peptide uncharacteristically long for

trypsin and not ideal for mass spectrometry. As such it is not considered as a surrogate peptide for P-III-NP LC/MS methods.

The activity of the *b*P-III-NP sample (~10 µg/mL) was assessed by immunoassay (2.7 µg/mL, 64.2 nM) using the UniQ® P-III-NP RIA, however the purity of the sample remains unknown. Data dependent analysis of peptide products obtained from trypsin digestion of this sample returned no hits for P-III-NP against the SwissProt database. This can be explained by the observation of PTMs affecting the observed mass of the *in silico*-derived peptides (not reflected in the database) discovered during the visual analysis of LC/HRAMS data using extracted ion chromatograms produced by Xcalibur.

Although helpful in identifying possible cleavage sites and presenting peptide fragments for MS analysis, the exclusion of PTM predictions in these search algorithms produced incomplete peptides for P-III-NP. Sole dependence on data obtained from these databases potentially prevents the identification of endogenous tryptic peptides without *a priori* knowledge of PTMs.

Identification of P-III-NP peptides from trypsin digested bovine material

Manual analysis of data obtained by LC/HRAMS on trypsin digested products of *b*P-III-NP, revealed a peptide eluting at 5.34 min with ions possessing a double charge at *m/z* at 1030.9400 (±5 ppm) and a triple charge at *m/z* 687.6291 (±5 ppm). A product ion scan spectrum for this peptide was generated from the doubly-charged ion, which when sequenced corresponds with the amino acid sequence Q_{-NH3}QEAVDGGCSHLGQSYADR. T1 originates from the Col 1 domain of P-III-NP, which is at the N-terminus of the protein. Post-translational cyclization of amino-terminal Q in proteins and peptides is achieved by de-amidation (loss of NH_3) to form a pyroglutamate (*p*E) residue, as is the case for P-III-NP. The absence of unmodified *b*T1 within the experimental trypsin peptide products suggests that circulating P-III-NP is completely deamidated at the N-terminal Q. This is not surprising as N-terminal pyroglutamation is mostly spontaneous, however, it is can also be catalysed by glutaminyl cyclase, an enzyme found in both plants and animals [21]. Cyclization of this N-terminus Q, protects P-III-NP from exopeptidase degradation and makes the protein more hydrophilic.

Another peptide eluting at 4.23 min, with ions possessing a single charge at *m/z* 1067.5116 (±5 ppm), double charge at *m/z* 534.2594 (±5 ppm) and triple charge at *m/z* 356.5087 (±5 ppm), was observed. A product ion scan spectrum was generated from the doubly-charged ion, which when sequenced corresponded to the amino acid sequence $GDP_{+OH}GPP_{+OH}GPGIP_{+OH}GPGR$. The addition of a hydroxyl group to proline in the Y position of the G-X-Y repeat is typical of collagen, where hydroxyproline residues at this site facilitate covalent hydrogen bonds to support the α-helical topology of this region. T5 originates from the Col 3 domain and so it is not surprising that hydroxyprolines are observed in this peptide.

No peptide corresponding to T6 was observed by LC/HRAMS, however, a potential N-glycosylation site is present at ^{127}N-Y-S-P. The attachment of a complex oligosaccharide

to P-III-NP has been previously observed [17], with the most likely site of attachment being *N*-127. Without prior removal or characterisation of this possible carbohydrate unit, if glycosylated, the mass of T6 generated by trypsin digestion of bovine P-III-NP is unknown. To overcome this, the carbohydrate moiety can be cleaved from the protein prior to digestion, however this step increases sample handling, may cause loss of yield and complicates sample preparation. Further scrutiny of the T6 peptide reveals a susceptible post-digest spontaneous asparagine de-amidation site at the N-terminal *N*-97. The non-enzymatic nature of this de-amidation leads to variability in the occurrence of this PTM [22], which affects the reproducibility of peptide generation *in situ.* Hence, T6 is not considered further as a potential surrogate for P-III-NP analysis.

Figure 1: Extracted ion chromatogram at *m/z* 1030.9400 ±5 ppm and corresponding product ion scan spectrum of the fragmented carbamidomethylated PTM *b*T1 triply-charged peptide (±1 u) using a normalised collision energy of 30 %. The most abundant b- and y-ions are annotated, where no charge is indicated the peptide is singly-charged.

Figure 2: Extracted ion chromatogram at *m/z* 534.2594 ±5 ppm and corresponding product ion scan spectrum of the fragmented PTM T5 doubly-charged peptide (±1 u) using a normalised collision energy of 20 %. The most abundant b- and y-ions are annotated, where no charge is indicated the peptide is singly-charged.

Peptide selection

Unique P-III-NP tryptic peptide T1 and T5 were fully characterised by MS-HRAMS analysis, with full coverage of the amino acid sequence through the identification of combined b- and y-ions produced by fragmentation. Confirmation of PTMs to the *in silico* sequence and the specificity of location is proven. Previous work with P-III-NP removed the pyroglutamate to facilitate Edman sequencing [11], data presented here shows evidence of the pyroglutamate replacing glutamine at the N-terminus of P-III-NP. For the T5 fragment, complete hydroxylation of proline in the Y position of the G-X-Y tri-peptide is shown; detailed analysis of full scan HRAMS data confirms the site specificity of this PTM, as no hP was observed for P in the X position.

Although P-III-NP characterisation was done with *b*P-III-NP, the *in silico* derived T5 peptide sequence for human is the same; hence the PTMs associated with this peptide should be conserved in humans. For the T1 peptide, *h*T1 has an aspartic acid residue at position 6 which is replaced by glutamic acid (E) in *b*T1, resulting in a decrease in molar mass of 15.023 Da. This mass shift facilitates the possible differentiation of species by MS. From previous experiments, we have observed deamidation of the N-terminus Q to pE in pure *in silico*-derived T1 peptide (QQEAVEGGCSHLGQSYADR), thus supporting the theory of spontaneous pyroglutamation. The cause of pyroglutamation of the P-III-NP in our bovine sample is unknown. Due to the inherent instability of the unmodified peptide, we suggest that synthesized T1 peptides should be produced with an N-terminus pyroglutamate.

Micro- and nano-flow LC/MS methods developed for PTM *h*T1 (*p*EQEAVDGGCSHLGQSYADR) had LODs of 100 pM (100 amoles) and 5 pM

(25 amoles), respectively, whilst for T5 (GD(*h*P)GP(*h*P)GI(*h*P)GR) LODs of 20 pM (20 amoles) and 5 pM (25 amoles) was achieved. Thus, both methods are suitable for the analysis of P-III-NP at basal serum concentrations (5 ng/mL, 125 pM). Chromatographic data generated from trypsin digested immuno-captured human serum P-III-NP show peaks corresponding to the expected PTM *h*T1 and T5 peptides. These results support the conservation of PTMs across bovine and human species, strengthening the use of our selected peptides (*h*T1 and T5) for the development of quantitative LC/MS methods targeting P-III-NP. Using our synthesised T5 peptide, we have semi-quantitatively estimated the basal concentration of *h*P-III-NP to be 50 pM (2 ng/mL) that are similar to concentrations estimated by immunoassay (Orion). To associate traceability to our methods, however, suitable reference material for the suggested peptides or intact P-III-NP is necessary.

Figure 3: Extracted ion chromatogram of P-III-NP peptides, *h*T1 (precursor *m/z* 1038 and 692) and T5 (precursor *m/z* 534), from a pooled human serum sample after trypsin digestion analysed by a) nano-flow LC/MS and b) micro-flow LC/MS. Peak intensity in relative abundance (%) and time (min) are shown on the y- and x-axes, respectively. The peak height (in absolute counts) and signal to noise ratio for each peak are annotated.

Conclusions

As a step towards developing a quantitative method based on mass spectrometry for analysing P-III-NP, the amino acid sequence of the protein has been partially sequenced. This has enabled the synthesis of appropriate peptides for the semi-quantitative LC/MS analysis of *h*P-III-NP in human serum.

P-III-NP peptides, T1 and T5, have been identified for the first time through the analysis of a trypsin digested *b*P-III-NP sample by LC/MS/HRAMS. Based on the structure of P-III-NP, PTMs can be predicted. However, protein algorithms do not include these events when *in silico*-digests are conducted. Our study highlights the importance of including PTMs for protein MS identification, as sole dependence on generated *in silico*-digest fragments can lead to the misidentification of peptide fragments. Characterisation and identification of human P-III-NP peptides *h*T1 and T5 brings us one step closer towards developing a MS quantification method for determining P-III-NP concentrations, in blood that will be adaptable to anti-doping and clinical testing and will complement current immunoassays. However, to develop commutable quantitative methods for these fields, standardised international reference for P-III-NP (as peptides or intact protein, the latter being preferred) as well as heavy labelled "internal standards" are needed.

Acknowledgement

The authors would like to thank Partnership for Clean Competition for funding the work presented; Institute for Bioanalytics, LLC for supplying P-III-NP standard used for analyses; and Mr. Christiaan Bartlett of the Drug Control Centre (King's College London) for technical assistance with immunoassay protocols.

References

[1] Jensen LT. The aminoterminal propeptide of type III procollagen - Studies on physiology and pathophysiology. Danish Medical Bulletin 1997;44(1):70-78.
[2] Rogol AD. Growth hormone and the adolescent athlete: What are the data for its safety and efficacy as an ergogenic agent? Growth Hormone & Igf Research 2009;19(4):294-299.
[3] Jeffers LJ, Coelholittle M, Cheinquer H, et al. Procollagen-III peptide and chronic viral C-hepatitis. American Journal of Gastroenterology 1995;90(9):1437-1440.
[4] Ramadori G, Zohrens G, Manns M, et al. Serum hyaluronate and type-III procollagen aminoterminal propeptide concentration in chronic liver-disease - relationship to cirrhosis and disease-activity. European Journal of Clinical Investigation 1991;21(3):323-330.
[5] Bentsen KD, Horslevpetersen K, Junker P, Juhl E, Lorenzen I. Serum aminoterminal procollagen type-III peptide in acute viral-hepatitis - a long-term follow-up-study. Liver 1987;7(2):96-105.
[6] Lichtinghagen R, Pietsch D, Bantel H, Manns, MP, Brand K, Bahr MJ. The Enhanced Liver Fibrosis (ELF) score: Normal values, influence factors and proposed cut-off values. Journal of Hepatology 2013;59(2):236-242..
[7] Trivedi P, Hindmarsh P, Risteli J, Risteli L, Mowat AP, Brook, CGD. Growth velocity, growth-hormone therapy, and serum concentrations of the amino-terminal propeptide of type-III procollagen. Journal of Pediatrics 198;114(2):225-230.
[8] Sönksen P. The International Olympic Committee (IOC) and GH-2000. Growth Hormone and IGF Research 2009;19(4):341-345
[9] Nelson AE, Ho KKY. Demographic factors influencing the GH system: Implications for the detection of GH doping in sport. Growth Hormone and IGF Research 2009;19(4):327-332.
[10] Nowack H, Olsen BR, Timpl R. Characterization of amino-terminal segment in type-3 procollagen. European Journal of Biochemistry 1976;70(1):205-216.
[11] Brandt A, Glanville RW, Horlein D, et al. Complete amino-acid-sequence of the N-terminal extension of calf skin type-III procollagen. Biochemical Journal 1984;219(2):625-634.
[12] Bruckner P, Bachinger HP, Timpl R, Engel J. 3 conformationally distinct domains in amino-terminal segment of type-III procollagen and its rapid triple helix reversible coil transition. European Journal of Biochemistry 1978;90(3):595-603.
[13] Knudsen CS, Heickendorff L, Nexo E. Measurement of amino terminal propeptide of type III procollagen (PIIINP) employing the ADVIA Centaur platform. Validation, reference interval and comparison to UniQ RIA. Clinical Chemistry and Laboratory Medicine 2014;52(2):237-241.
[14] Powrie JK, Bassett EE, Rosen T, et al. Detection of growth hormone abuse in sport, Growth Hormone and IGF Research 2007;17(3):220-226.
[15] Cowan DA, Bartlett C. Laboratory issues in the implementation of the marker method. Growth Hormone and IGF Research 2009;19(4):357-360.
[16] Tabb DL, Huang YY, Wysocki VH, Yates JR. Influence of basic residue content on fragment ion peak intensities in low-energy - collision-induced dissociation spectra of peptide. Analytical Chemistry 2004;76(5):1243-1248.
[17] Shinkai H, Lapiere CM. Characterization of oligosaccharide units of p-n-collagen type-III from dermatosparactic bovine skin. Biochimica et Biophysica Acta 1983;758(1):30-36.
[18] *UniProt. UniProt: the universal protein knowledgebase. Nucleic Acids Research 2017;45(D1):D158-D169.*

[19] Gasteiger E, Hoogland C, Gattiker A, et al. Protein Identification and Analysis Tools on the ExPASy Server, In: Walker J(ed). The Proteomics Protocols Handbook: Humana Press; 2005: 571-607.
[20] U.S.C.F. Mass Spectrometry Facility, 5.12.2 ed. The Regents of the University of California. The University of California **1996-2014**.
[21] Liu YD, Goetze AM, Bass RB, Flynn GC. N-terminal Glutamate to Pyroglutamate Conversion in Vivo for Human IgG2 Antibodies. Journal of Biological Chemistry 2011;286(13):11211-11217.
[22] Krokhin OV, Antonovici M, Ens W, Wilkins JA, Standing KG. Deamidation of -Asn-Gly-sequences during sample preparation for proteomics: Consequences for MALDI and HPLC-MALDI analysis. Analytical Chemistry 2006;78(18):6645-6650.

P-III-NP Precursor Peptide	Precursor ion (*m/z*)	Product ion (*m/z*)	Collision energy (eV)
***h*T1**	692 $[M+3H]^{3+}$	768 (b_8)	10
	1038 $[M+2H]^{2+}$	240 (b_2)	50
	1038 $[M+2H]^{2+}$	440 (b_4)	40
T5	534 $[M+2H]^{2+}$	363 (y_7^{2+})	20
	534 $[M+2H]^{2+}$	448 (y_9^{2+})	13
	534 $[M+2H]^{2+}$	628 (y_6)	27

Table 1: Collision energies for SRM (selected reaction monitoring) transitions acquired for P-III-NP peptides by micro- and nano-ESI MS.

Peptide Reference	Site of peptide in P-III-NP (aa)	Peptide sequence	Peptide length (aa)	Molar Mass (Da)
***h*T1**	1-19	QQEAV**E**GGCSHLGQSYADR	19	2034.893
***b*T1**	1-19	QQEAV**D**GGCSHLGQSYADR	19	2019.870
T2-4	20-85	DVWKPEPCQICVCDSGSVLCDDII CDDQELDCPNPEIPFGECCAVCPQ PPTAPTRPPNGQGPQGPK	66	7020.075
T5	86-96	GDPGPPGIPGR	11	1019.527
***h*T6**	97-130	NGDPG**I**PG**Q**PGSPGSPGPPGICESC PTG**P**QNYSP	34	3218.415
***b*T6**	97-130	NGDPG**P**PG**S**PGSPGSPGPPGICESC PTG**G**QNYSP	34	3120.319

Table 2: Unique trypsin *in silico*-derived digestion products for human (UniProt accession no. P02461(24-153)) and bovine (UniProt accession no. Q08E14 (24-153)) P-III-NP with no missed cleavages allowed. Where the amino acid sequence of peptides differ between species (residues highlighted in bold), *h* and *b,* represent human and bovine respectively.

Peptide	Amino acid sequence	Peptide ion (*m/z*)		
		$[M+H]^+$	$[M+2H]^{2+}$	$[M+3H]^{3+}$
In silico T1	QQEAVDGGCSHLGQSYADR	2020.8777	1010.9425	674.2974
Observed carbamidomethylated PTM T1	*p*EQEAVDGGCSHLGQSYADR	2060.8727	1030.9400	687.6291
In silico T5	GDPGPPGIPGR	1019.5269	510.2671	340.5138
Observed PTM T5	GD(*h*P)GP(*h*P)GI(*h*P)GR	1067.5116	534.2594	356.5087
In silico T6	NGDPGPPGSPGSPGSPGPPGICESCPTGGQNYSP	3121.3262	1561.1667	1041.1136

Table 3: Bovine P-III-NP trypsin digestion products selected for MS analysis, peptides shown are the theoretically derived *in silico* (Protein Cutter [19] and Protein Prospector [20]) and those observed with a PTM form within the experimental digest product. Pyroglutamate and hydroxylated proline residues are represented by *p*E and *h*P, respectively. Singly- ($[M+H]^+$), doubly- ($[M+2H]^{2+}$) and triply- ($[M+3H]^{3+}$) charged ions for each peptide are presented.

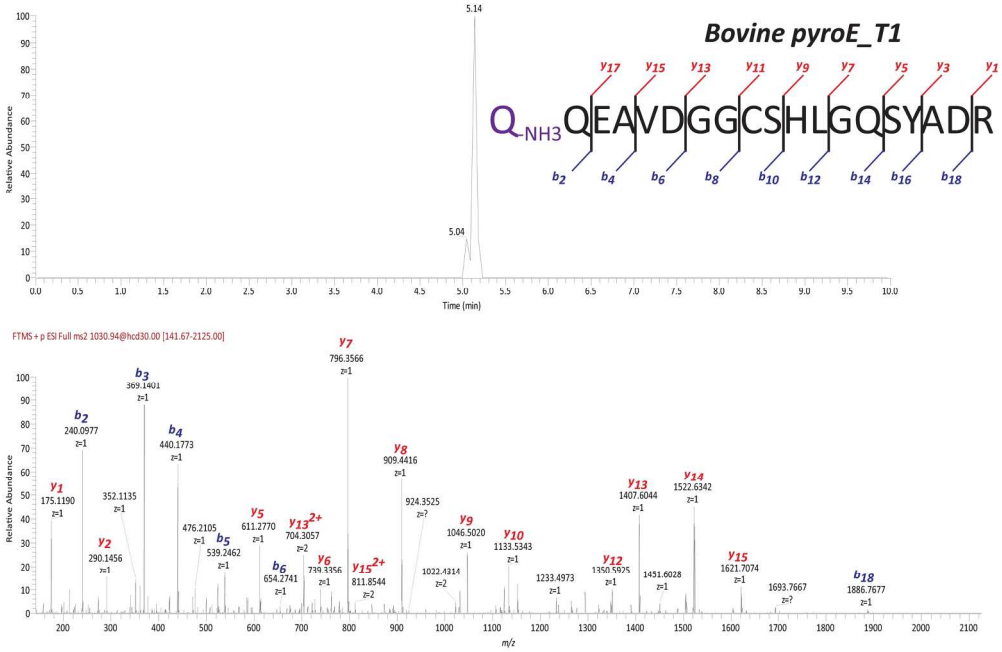

Figure 1: Extracted ion chromatogram at m/z 1030.9400 ±5 ppm and corresponding product ion scan spectrum of the carbamidomethylated PTM *b*T1 triply-charged peptide (±1 u) using a normalised collision energy of 30 %. The most abundant b- and y-ions are annotated, where no charge is indicated the peptide is singly-charged.

196x130mm (300 x 300 DPI)

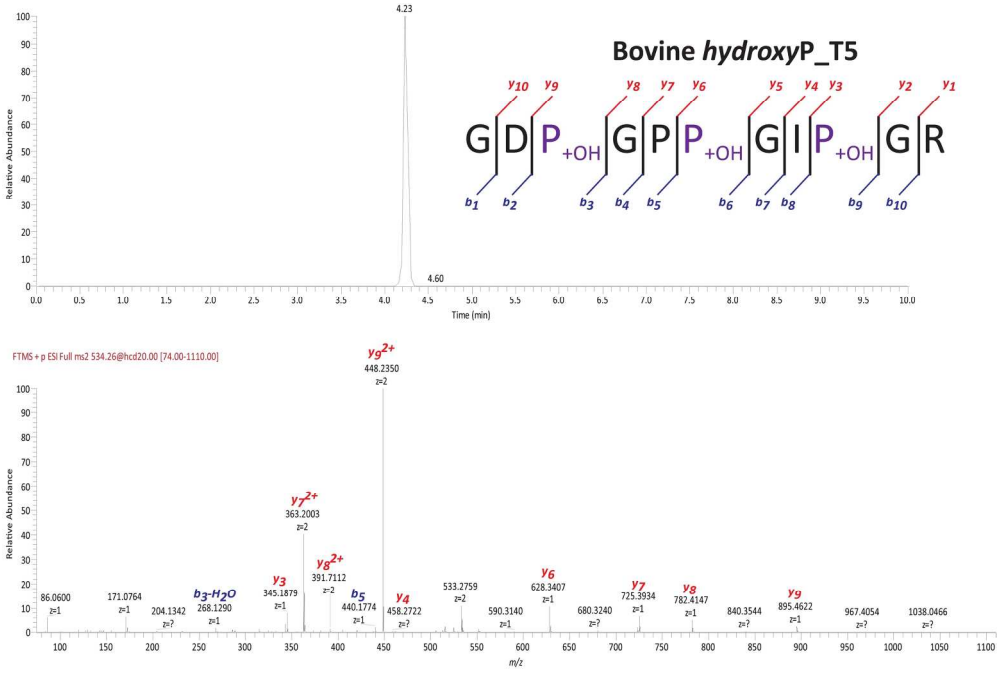

Figure 2: Extracted ion chromatogram at m/z 534.2594 ±5 ppm and corresponding product ion scan spectrum of the fragmented PTM T5 doubly-charged peptide (±1 u) using a normalised collision energy of 20 %. The most abundant b- and y-ions are annotated, where no charge is indicated the peptide is singly-charged.

197x134mm (300 x 300 DPI)

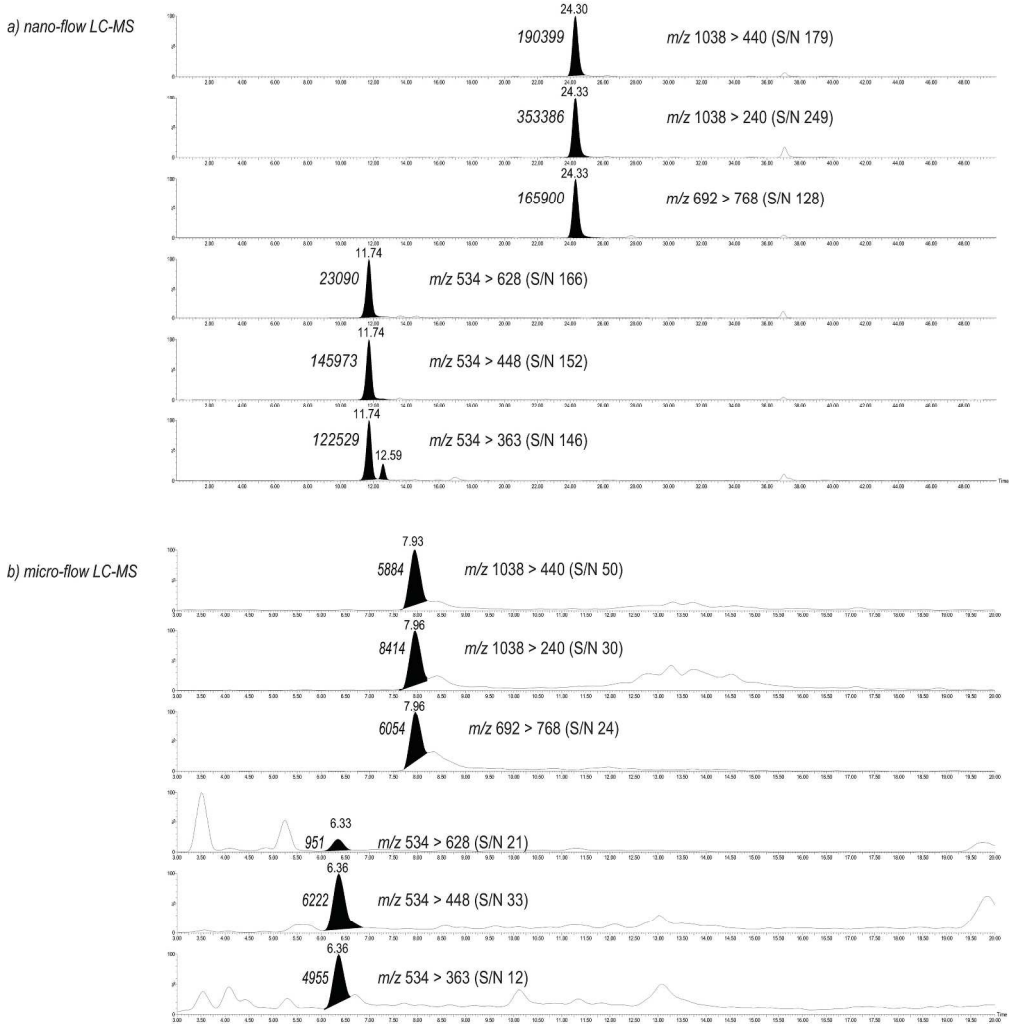

Figure 3: Extracted ion chromatogram of P-III-NP peptides, *h*T1 (precursor m/z 1038 and 692) and T5 (precursor m/z 534), from a pooled human serum sample after trypsin digestion analysed by a) nano-flow LC/MS and b) micro-flow LC/MS. Peak intensity in relative abundance (%) and time (min) are shown on the y- and x-axes, respectively. The peak height (in absolute counts) and signal to noise ratio for each peak are annotated.

249x256mm (300 x 300 DPI)